

SEQUENCE PROTOCOL

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Prof.DR. Werner Lubitz
- (B) ROAD: Schoenborngasse 12/7
- (C) CITY: Vienna
- (E) COUNTRY: Austria
- (F) POSTAL CODE: 1080

(ii) TITLE OF INVENTION: New systems for the regulation
of gene expression

(iii) NUMBER OF SEQUENCES: 10

(iv) COMPUTER-READABLE FORM:

- (A) DATA CARRIER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0,
Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(vi) INITIAL ORIGIN:

- (A) ORGANISM: lambda OR operator (wild-type)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACGTTAAATC TATCACCGCA AGGGATAAAT ATCTAACACC GTGCGTGTTG ACTATTTTAC 60
CTCTGGCGGT GATAATGGTT GC 82

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(vi) INITIAL ORIGIN:

(A) ORGANISM: lambda OR operator (mutant)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ACGTTAAATC TATCACGCA AGGGATAAAT ATCTAACACC GCGCGTGTG ACTATTTTAC
CTCTGGCGGT GATAATGGTT GC

60

82

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(vi) INITIAL ORIGIN:

(A) ORGANISM: lambda OL operator (wild-type)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ACATACAGAT AACCATCTGC GGTGATAAAT TATCTCTGGC GGTGTTGACA TAAATACCAC
TGGCGGTGAT ACTGAGCACA TCAGC

60

85

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1601 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: double strand
- (D) TOPOLOGY: both

(vi) INITIAL ORIGIN:

(A) ORGANISM: pAW12 fragment

(xi) CHARACTERISTICS:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (106..816)

*Sub
a
cont*

(ix) CHARACTERISTICS:

(A) NAME/KEY: CDS

(B) LOCATION: 1144..1416

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

10

ATTTACTATG TTATGTTCTG AGGGGAGTGA AAATTCCCCT AATTCGATGA AGATTCTTGC 60
TCAATTGTGA TCAGCTATGC GCCGACCAGA ACACCTTGCC GATCAGCCAA ACGTCTCTTC 120
AGGCCACTGA CTAGCGATAA CTTTCCCAC AACGGAACAA CTCTCATTCG ATGGGATCAT 180
TGGGTACTGT GGGTTTAGTG GTTGTAATAA CACCTGACCG CTATCCCTGA TCAGTTTCTT 240
GAAGGTAAAC TCATCACCCC CAAGTCTGGC TATGCAGAAA TCACCTGGCT CAACAGCCTG 300
CTCAGGGTCA ACGAGAATTA ACATTCCGTG AGGAAAGCTT GGCTTGGAGC CTGTTGGTGC 360
GGTCATGGAA TTACCTTCAA CCTCAAGCCA GAATGCAGAA TCACTGGCTT TTTTGGTTGT 420
GCTTACCCAT CTCTCCGCAT CACCTTTGGT AAAGGTTCTA AGCTTAGGTG AGAACATCCC 480
TGCCTGAACA TGAGAAAAAA CAGGGTACTC ATACTCACTT CTAAGTGACG GCTGCATACT 540
AACCCTTCA TACATCTCGT AGATTCTCTT GGCGATTGAA GGGCTAAATT CTTCAACGCT 600
AACTTTGAGA ATTTTTGTAA GCAATGCGGC GTTATAAGCA TTTAATGCAT TGATGCCATT 660
AAATAAGCA CCAACGCCTG ACTGCCCCAT CCCCATCTTG TCTGCGACAG ATTCCTGGGA 720
TAAGCCAAGT TCATTTTTCT TTTTTCATA AATTGOTTTA AGGCGACGTG CGTCCTCAAG 780
CTGCTCTTGT GTTAATGGTT TCTTTTTTGT GCTCATACGT TAAATCTATC ACCGCAAGGG 840
ATAAATATCT AACACCGCGC GTGTTGACTA TTTTACCTCT GGCGGTGATA ATGGTTGCAT 900
GTACTAAGTA GGTGTATGG AACAAACGCAT AACCCTGAAA GATTATGCAA TGCCTTTGG 960
GCAAACCAAG ACAGCTAAAG ATCCTCTAGA GTCGACCTGC AGGCATGCAA GCTTATCGAA 1020
TTCTCATTCA GGCTTCTGCC GTTTTGGATT TAACCGAAGA TGATTTTCGAT TTTCTGACGA 1080
GTAACAAAGT TTGGATTGCT ACTGACCGCT CTCGTGCTCG TCGCTGCGTT GAGGCTTGCG 1140
TTT ATG GTA CGC TGG ACT TTG TGG GAT ACC CTC GCT TTC CTG CTC CTG 1188
Met Val Arg Trp Thr Leu Trp Asp Thr Leu Ala Phe Leu Leu Leu
1 5 10 15
TTG AGT TTA TTG CTG CCG TCA TTG CTT ATT ATG TTC ATC CCG TCA ACA 1236
Leu Ser Leu Leu Leu Pro Ser Leu Leu Ile Met Phe Ile Pro Ser Thr
20 25 30
TTC AAA CGG CCT GTC TCA TCA TGG AAG GCG CTG AAT TTA CGG AAA ACA 1284
Phe Lys Arg Pro Val Ser Ser Trp Lys Ala Leu Asn Leu Arg Lys Thr
35 40 45

65

TTA TTA ATG GCG TCG AGC GTC CGG TTA AAG CCG CTG AAT TGT TCG CGT 1332
 Leu Leu Met Ala Ser Ser Val Arg Leu Lys Pro Leu Asn Cys Ser Arg
 50 55 60

5 TTA CCT TGC GTG TAC GCG CAG GAA ACA CTG ACG TTC TTA CTG ACG CAG 1380
 Leu Pro Cys Val Tyr Ala Gln Glu Thr Leu Thr Phe Leu Leu Thr Gln
 65 70 75

AAG AAA ACG TGC GTC AAA AAT TAC GTG CAG AAG GAG TGATGTAATG 1426
 10 Lys Lys Thr Cys Val Lys Asn Tyr Val Gln Lys Glu
 80 85 90

TCTAAAGGTA AAAACGTTT TGGCGCTCGC CCTGGTCGTC CGCAGCCGTT GCGAGGTACT 1486

15 AAAGGCAAGC GTAAAGGCGC TCGTCTTTGG TATGTAGGTG GTCAACAATT TTAATTGCAG 1546

GGGCTTCGGC CCTTACTTGA GGATAAATTA TGTCTAATAT TCAAAGTGGC GCCGA 1601

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 237 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

35 Met Ser Thr Lys Lys Lys Pro Leu Thr Gln Glu Gln Leu Glu Asp Ala 15
 1 5 10

Arg Arg Leu Lys Ala Ile Tyr Glu Lys Lys Lys Asn Glu Leu Gly Leu 30
 20 25 30

40 Ser Gln Glu Ser Val Ala Asp Lys Met Gly Met Gly Gln Ser Gly Val 45
 35 40 45

45 Gly Ala Leu Phe Asn Gly Ile Asn Ala Leu Asn Ala Tyr Asn Ala Ala 60
 50 55 60

Leu Leu Thr Lys Ile Leu Lys Val Ser Val Glu Glu Phe Ser Pro Ser 80
 65 70 75 80

50 Ile Ala Arg Glu Ile Tyr Glu Met Tyr Glu Ala Val Ser Met Gln Pro 95
 85 90 95

Ser Leu Arg Ser Glu Tyr Glu Tyr Pro Val Phe Ser His Val Gln Ala 110
 100 105 110

55 Gly Met Phe Ser Pro Lys Leu Arg Thr Phe Thr Lys Gly Asp Ala Glu 125
 115 120 125

60 Arg Trp Val Ser Thr Thr Lys Lys Ala Ser Asp Ser Ala Phe Trp Leu 140
 130 135 140

Glu Val Glu Gly Asn Ser Met Thr Ala Pro Thr Gly Ser Lys Pro Ser 160
 145 150 155 160

See a. cont.

Phe Pro Asp Gly Met L u Ile L u Val Asp Pro Glu Gln Ala Val Glu
165 170 175
5 Pro Gly Asp Phe Cys Ile Ala Arg Leu Gly Gly Asp Glu Phe Thr Phe
180 185 190
Lys Lys Leu Ile Arg Asp Ser Gly Gln Val Phe Leu Gln Pro Leu Asn
195 200 205
10 Pro Gln Tyr Pro Met Ile Pro Cys Asn Glu Ser Cys Ser Val Val Gly
210 215 220
Lys Val Ile Ala Ser Gln Trp Pro Glu Glu Thr Phe Gly
225 230 235
15

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 91 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

- 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

30 Met Val Arg Trp Thr Leu Trp Asp Thr Leu Ala Phe Leu Leu Leu Leu
1 5 10 15
35 Ser Leu Leu Leu Pro Ser Leu Leu Ile Met Phe Ile Pro Ser Thr Phe
20 25 30
Lys Arg Pro Val Ser Ser Trp Lys Ala Leu Asn Leu Arg Lys Thr Leu
35 40 45
40 Leu Met Ala Ser Ser Val Arg Leu Lys Pro Leu Asn Cys Ser Arg Leu
50 55 60
45 Pro Cys Val Tyr Ala Gln Glu Thr Leu Thr Phe Leu Leu Thr Gln Lys
65 70 75 80
Lys Thr Cys Val Lys Asn Tyr Val Gln Lys Glu
85 90

50

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 2834 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: both

60

Sub
a'
cont

(vi) INITIAL ORIGIN:
(A) ORGANISM: pCSJ fragment

(xi) CHARACTERISTICS:
(A) NAME/KEY: CDS
(B) LOCATION: complement (106..816)

(ix) CHARACTERISTICS:
(A) NAME/KEY: CDS
(B) LOCATION: 1025..2104

(ix) CHARACTERISTICS:
(A) NAME/KEY: CDS
(B) LOCATION: 2377..2649

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

20

ATTTACTATG TTATGTTCTG AGGGGAGTGA AAATTCCCCT AATTCGATGA AGATTCTTGC 60

25 TCAATTGTTA TCAGCTATGC GCCGACCAGA ACACCTTGCC GATCAGCCAA ACGTCTCTTC 120

AGGCCACTGA CTAGCGATAA CTTTCCCAC AACGGAACAA CTCTCATTGC ATGGGATCAT 180

30 TGGGTACTGT GGGTTTAGTG GTTGTA AAAA CACCTGACCG CTATCCCTGA TCAGTTTCTT 240

GAAGGTAAAC TCATCACCCC CAAGTCTGGC TATGCAGAAA TCACCTGGCT CAACAGCCTG 300

CTCAGGGTCA ACGAGAATTA ACATTCCGTC AGGAAAGCTT GGCTTGGAGC CTGTTGGTGC 360

35 GGTCATGGAA TTACCTTCAA CCTCAAGCCA GAATGCAGAA TCACTGGCTT TTTTGGTTGT 420

GCTTACCCAT CTCTCCGCAT CACCTTTGGT AAAGGTTCTA AGCTTAGGTG AGAACATCCC 480

40 TGCCTGAACA TGAGAAAAAA CAGGGTACTC ATACTCACTT CTAAGTGACG GCTGCATACT 540

AACCGCTTCA TACATCTCGT AGATTTCTCT GGCGATTGAA GGGCTAAATT CTTCAACGCT 600

AACTTTGAGA ATTTTGTAA GCAATGCGGC GTTATAAGCA TTTAATGCAT TGATGCCATT 660

45 AAATAAAGCA CCAACGCCCTG ACTGCCCCAT CCCCATCTTG TCTGCGACAG ATTCCTGGGA 720

TAAGCCAAGT TCATTTTTCT TTTTTCATA AATTGCTTTA AGGCGACGTG CGTCCTCAAG 780

50 CTGCTCTTGT GTTAATGGTT TCTTTTTTGT GCTCATACGT TAAATCTATC ACCGCAAGGG 840

ATAAATATCT AACACCGCGC GTGTTGACTA TTTTACCTCT GGCGGTGATA ATGGTTGCAT 900

GTAATAAGTA GGTGTATGG AACAACGCAT AACCCTGAAA GATTATGCAA TGCCTTTGG 960

55 GCAAACCAAG ACAGCTAAAG ATCCTCTAGA GCGCCCGGAA GAGAGTCAAT TCAGGGTGGT 1020

GAAT GTG AAA CCA GTA ACG TTA TAC GAT GTC GCA GAG TAT GCC GGT GTC 1069

Val Lys Pro Val Thr Leu Tyr Asp Val Ala Glu Tyr Ala Gly Val
95 100 105

60

TCT TAT CAG ACC GTT TCC CGC GTG GTG AAC CAG GCG AGC CAC GTT TCT 1117

Ser Tyr Gln Thr Val Ser Arg Val Val Asn Gln Ala Ser His Val Ser
110 115 120

Sub
a1
cont

	GCG	AAA	ACG	CGG	GAA	AAA	GTG	GAA	GCG	GCG	ATG	GCG	GAG	CTG	AAT	TAC	1165
	Ala	Lys	Thr	Arg	Glu	Lys	Val	Glu	Ala	Ala	Met	Ala	Glu	Leu	Asn	Tyr	
			125					130					135				
5	ATT	CCC	AAC	CGC	GTG	GCA	CAA	CAA	CTG	GCG	GGC	AAA	CAG	TCG	TTG	CTG	1213
	Ile	Pro	Asn	Arg	Val	Ala	Gln	Gln	Leu	Ala	Gly	Lys	Gln	Ser	Leu	Leu	
		140					145					150					
10	ATT	GGC	GTT	GCC	ACC	TCC	AGT	CTG	GCC	CTG	CAC	GCG	CCG	TCG	CAA	ATT	1261
	Ile	Gly	Val	Ala	Thr	Ser	Ser	Leu	Ala	Leu	His	Ala	Pro	Ser	Gln	Ile	
		155				160					165					170	
15	GTC	GCG	GCG	ATT	AAA	TCT	CGC	GCC	GAT	CAA	CTG	GGT	GCC	AGC	GTG	GTG	1309
	Val	Ala	Ala	Ile	Lys	Ser	Arg	Ala	Asp	Gln	Leu	Gly	Ala	Ser	Val	Val	
					175					180					185		
20	GTG	TCG	ATG	GTA	GAA	CGA	AGC	GGC	GTC	GAA	GCC	TGT	AAA	GCG	GCG	GTG	1357
	Val	Ser	Met	Val	Glu	Arg	Ser	Gly	Val	Glu	Ala	Cys	Lys	Ala	Ala	Val	
				190					195					200			
25	CAC	AAT	CTT	CTC	GCG	CAA	CGC	GTC	AGT	GGG	CTG	ATC	ATT	AAC	TAT	CCG	1405
	His	Asn	Leu	Ala	Gln	Arg	Val	Ser	Gly	Leu	Ile	Ile	Asn	Tyr	Pro		
			205				210					215					
30	CTG	GAT	GAC	CAG	GAT	GCC	ATT	GCT	GTG	GAA	GCT	GCC	TGC	ACT	AAT	GTT	1453
	Leu	Asp	Asp	Gln	Asp	Ala	Ile	Ala	Val	Glu	Ala	Ala	Cys	Thr	Asn	Val	
		220					225					230					
35	CCG	GCG	TTA	TTT	CTT	GAT	GTC	TCT	GAC	CAG	ACA	CCC	ATC	AAC	AGT	ATT	1501
	Pro	Ala	Leu	Phe	Leu	Asp	Val	Ser	Asp	Gln	Thr	Pro	Ile	Asn	Ser	Ile	
		235				240					245					250	
40	ATT	TTC	TCC	CAT	GAA	GAC	GGT	ACG	CGA	CTG	GGC	GTG	GAG	CAT	CTG	GTC	1549
	Ile	Phe	Ser	His	Glu	Asp	Gly	Thr	Arg	Leu	Gly	Val	Glu	His	Leu	Val	
					255					260					265		
45	GCA	TTG	GGT	CAC	CAG	CAA	ATC	GCG	CTG	TTA	GCG	GGC	CCA	TTA	AGT	TCT	1597
	Ala	Leu	Gly	His	Gln	Gln	Ile	Ala	Leu	Leu	Ala	Gly	Pro	Leu	Ser	Ser	
				270					275					280			
50	GTC	TCG	GCG	CGT	CTG	CGT	CTG	GCT	GGC	TGG	CAT	AAA	TAT	CTC	ACT	CGC	1645
	Val	Ser	Ala	Arg	Leu	Arg	Leu	Ala	Gly	Trp	His	Lys	Tyr	Leu	Thr	Arg	
			285				290						295				
55	AAT	CAA	ATT	CAG	CCG	ATA	GCG	GAA	CGG	GAA	GGC	GAC	TGG	AGT	GCC	ATG	1693
	Asn	Gln	Ile	Gln	Pro	Ile	Ala	Glu	Arg	Glu	Gly	Asp	Trp	Ser	Ala	Met	
		300					305					310					
60	TCC	GGT	TTT	CAA	CAA	ACC	ATG	CAA	ATG	CTG	AAT	GAG	GGC	ATC	GTT	CCC	1741
	Ser	Gly	Phe	Gln	Gln	Thr	Met	Gln	Met	Leu	Asn	Glu	Gly	Ile	Val	Pro	
		315				320					325					330	
65	ACT	GCG	ATG	CTG	GTT	GCC	AAC	GAT	CAG	ATG	GCG	CTG	GGC	GCA	ATG	CGC	1789
	Thr	Ala	Met	Leu	Val	Ala	Asn	Asp	Gln	Met	Ala	Leu	Gly	Ala	Met	Arg	
					335					340					345		
70	GCC	ATT	ACC	GAG	TCC	GGG	CTG	CGC	GTT	GGT	GCG	GAT	ATC	TCG	GTA	GTG	1837
	Ala	Ile	Thr	Glu	Ser	Gly	Leu	Arg	Val	Gly	Ala	Asp	Ile	Ser	Val	Val	
				350					355					360			
75	GGA	TAC	GAC	GAT	ACC	GAA	GAC	AGC	TCA	TGT	TAT	ATC	CCG	CCG	TCA	ACC	1885
	Gly	Tyr	Asp	Asp	Thr	Glu	Asp	Ser	Ser	Cys	Tyr	Ile	Pro	Pro	Ser	Thr	
			365				370						375				
80	ACC	ATC	AAA	CAG	GAT	TTT	CGC	CTG	CTG	GGG	CAA	ACC	AGC	GTG	GAC	CGC	1933
	Thr	Ile	Lys	Gln	Asp	Phe	Arg	Leu	Leu	Gly	Gln	Thr	Ser	Val	Asp	Arg	
		380					385					390					

Sub
a:
cont.

	TTG CTG CAA CTC TCT CAG GGC CAG GCG GTG AAG GGC AAT CAG CTG TTG	1981
	Leu Leu Gln Leu Ser Gln Gly Gln Ala Val Lys Gly Asn Gln Leu Leu	
	395 400 405 410	
5	CCC GTC TCA CTG GTG AAA AGA AAA ACC ACC CTG GCG CCC AAT ACG CAA	2029
	Pro Val Ser Leu Val Lys Arg Lys Thr Thr Leu Ala Pro Asn Thr Gln	
	415 420 425	
10	ACC GCC TCT CCC CGC GCG TTG GCC GAT TCA TTA ATG CAG CTG GCA CGA	2077
	Thr Ala Ser Pro Arg Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg	
	430 435 440	
15	CAG GTT TCC CGA CTG GAA AGC GGG CAG TGAGCGCAAC GCAATTAATG	2124
	Gln Val Ser Arg Leu Glu Ser Gly Gln	
	445 450	
	TGAGTTAGCT CACTCATTAG GCACCCCAAG CTTTACACTT TATGCTTCCG GCTCGTATGT	2184
20	TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTCTGCA GGCATGCAAG	2244
	CTTATCGAAT TCTCATTCA GCTTCTGCCG TTTTGGATTT AACCGAAGAT GATTTTCGATT	2304
	TTCTGACGAG TAACAAAGTT TGGATTGCTA CTGACCGCTC TCGTGCTCGT CGCTGCGTTG	2364
25	AGGCTTGCGT TT ATG GTA CGC TGG ACT TTG TGG GAT ACC CTC GCT TTC	2412
	Met Val Arg Trp Thr Leu Trp Asp Thr Leu Ala Phe	
	1 5 10	
30	CTG CTC CTG TTG AGT TTA TTG CTG CCG TCA TTG CTT ATT ATG TTC ATC	2460
	Leu Leu Leu Leu Ser Leu Leu Leu Pro Ser Leu Leu Ile Met Phe Ile	
	15 20 25	
35	CCG TCA ACA TTC AAA CGG CCT GTC TCA TCA TGG AAG GCG CTG AAT TTA	2508
	Pro Ser Thr Phe Lys Arg Pro Val Ser Ser Trp Lys Ala Leu Asn Leu	
	30 35 40	
40	CGG AAA ACA TTA TTA ATG GCG TCG AGC GTC CGG TTA AAG CCG CTG AAT	2556
	Arg Lys Thr Leu Leu Met Ala Ser Ser Val Arg Leu Lys Pro Leu Asn	
	45 50 55 60	
	TGT TCG CGT TTA CCT TGC GTG TAC GCG CAG GAA ACA CTG ACG TTC TTA	2604
	Cys Ser Arg Leu Pro Cys Val Tyr Ala Gln Glu Thr Leu Thr Phe Leu	
	65 70 75	
45	CTG ACG CAG AAG AAA ACG TGC GTC AAA AAT TAC GTG CAG AAG GAG	2649
	Leu Thr Gln Lys Lys Thr Cys Val Lys Asn Tyr Val Gln Lys Glu	
	80 85 90	
50	TGATGTAATG TCTAAAGGTA AAAAACGTTT TGGCGCTCGC CCTGGTCGTC CGCAGCCGTT	2709
	GCGAGGTACT AAAGGCAAGC GTAAAGGCGC TCGTCTTTGG TATGTAGGTG GTCAACAATT	2769
	TTAATTGCAG GGGCTTCGGC CCTTACTTGA GGATAAATTA TGTCTAATAT TCAAACCTGGC	2829
55	GCCGA	2834

60 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

65

Sub
a1
Cont

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

5 Met Ser Thr Lys Lys Lys Pro Leu Thr Gln Glu Gln Leu Glu Asp Ala
1 5 10 15
Arg Arg Leu Lys Ala Ile Tyr Glu Lys Lys Lys Asn Glu Leu Gly Leu
20 25 30
10 Ser Gln Glu Ser Val Ala Asp Lys Met Gly Met Gly Gln Ser Gly Val
35 40 45
15 Gly Ala Leu Phe Asn Gly Ile Asn Ala Leu Asn Ala Tyr Asn Ala Ala
50 55 60
Leu Leu Thr Lys Ile Leu Lys Val Ser Val Glu Glu Phe Ser Pro Ser
65 70 75 80
20 Ile Ala Arg Glu Ile Tyr Glu Met Tyr Glu Ala Val Ser Met Gln Pro
85 90 95
Ser Leu Arg Ser Glu Tyr Glu Tyr Pro Val Phe Ser His Val Gln Ala
100 105 110
25 Gly Met Phe Ser Pro Lys Leu Arg Thr Phe Thr Lys Gly Asp Ala Glu
115 120 125
Arg Trp Val Ser Thr Thr Lys Lys Ala Ser Asp Ser Ala Phe Trp Leu
130 135 140
Glu Val Glu Gly Asn Ser Met Thr Ala Pro Thr Gly Ser Lys Pro Ser
145 150 155 160
35 Phe Pro Asp Gly Met Leu Ile Leu Val Asp Pro Glu Gln Ala Val Glu
165 170 175
Pro Gly Asp Phe Cys Ile Ala Arg Leu Gly Gly Asp Glu Phe Thr Phe
180 185 190
40 Lys Lys Leu Ile Arg Asp Ser Gly Gln Val Phe Leu Gln Pro Leu Asn
195 200 205
Pro Gln Tyr Pro Met Ile Pro Cys Asn Glu Ser Cys Ser Val Val Gly
210 215 220
45 Lys Val Ile Ala Ser Gln Trp Pro Glu Glu Thr Phe Gly
225 230 235
50

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

*Sub
at
con*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 Val Lys Pro Val Thr Leu Tyr Asp Val Ala Glu Tyr Ala Gly Val Ser
 1 5 10 15
 Tyr Gln Thr Val Ser Arg Val Val Asn Gln Ala Ser His Val Ser Ala
 20 25 30
 10 Lys Thr Arg Glu Lys Val Glu Ala Ala Met Ala Glu Leu Asn Tyr Ile
 35 40 45
 Pro Asn Arg Val Ala Gln Gln Leu Ala Gly Lys Gln Ser Leu Leu Ile
 50 55 60
 15 Gly Val Ala Thr Ser Ser Leu Ala Leu His Ala Pro Ser Gln Ile Val
 65 70 75 80
 Ala Ala Ile Lys Ser Arg Ala Asp Gln Leu Gly Ala Ser Val Val Val
 85 90 95
 Ser Met Val Glu Arg Ser Gly Val Glu Ala Cys Lys Ala Ala Val His
 100 105 110
 25 Asn Leu Leu Ala Gln Arg Val Ser Gly Leu Ile Ile Asn Tyr Pro Leu
 115 120 125
 Asp Asp Gln Asp Ala Ile Ala Val Glu Ala Ala Cys Thr Asn Val Pro
 130 135 140
 30 Ala Leu Phe Leu Asp Val Ser Asp Gln Thr Pro Ile Asn Ser Ile Ile
 145 150 155 160
 Phe Ser His Glu Asp Gly Thr Arg Leu Gly Val Glu His Leu Val Ala
 165 170 175
 35 Leu Gly His Gln Gln Ile Ala Leu Leu Ala Gly Pro Leu Ser Ser Val
 180 185 190
 40 Ser Ala Arg Leu Arg Leu Ala Gly Trp His Lys Tyr Leu Thr Arg Asn
 195 200 205
 Gln Ile Gln Pro Ile Ala Glu Arg Glu Gly Asp Trp Ser Ala Met Ser
 210 215 220
 45 Gly Phe Gln Gln Thr Met Gln Met Leu Asn Glu Gly Ile Val Pro Thr
 225 230 235 240
 Ala Met Leu Val Ala Asn Asp Gln Met Ala Leu Gly Ala Met Arg Ala
 245 250 255
 50 Ile Thr Glu Ser Gly Leu Arg Val Gly Ala Asp Ile Ser Val Val Gly
 260 265 270
 55 Tyr Asp Asp Thr Glu Asp Ser Ser Cys Tyr Ile Pro Pro Ser Thr Thr
 275 280 285
 Ile Lys Gln Asp Phe Arg Leu Leu Gly Gln Thr Ser Val Asp Arg Leu
 290 295 300
 60 Leu Gln Leu Ser Gln Gly Gln Ala Val Lys Gly Asn Gln Leu Leu Pro
 305 310 315 320
 65 Val Ser Leu Val Lys Arg Lys Thr Thr Leu Ala Pro Asn Thr Gln Thr
 325 330 335

Sub
A:
cont.

Ala Ser Pro Arg Ala Leu Ala Asp Ser Leu Met Gln Leu Ala Arg Gln
340 345 350

5 Val Ser Arg Leu Glu S r Gly Gln
355 360

10 (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) TYPE OF MOLECULE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

20 Met Val Arg Trp Thr Leu Trp Asp Thr Leu Ala Phe Leu Leu Leu Leu
1 5 10 15
25 Ser Leu Leu Leu Pro Ser Leu Leu Ile Met Phe Ile Pro Ser Thr Phe
20 25 30
30 Lys Arg Pro Val Ser Ser Trp Lys Ala Leu Asn Leu Arg Lys Thr Leu
35 40 45
Leu Met Ala Ser Ser Val Arg Leu Lys Pro Leu Asn Cys Ser Arg Leu
50 55 60
35 Pro Cys Val Tyr Ala Gln Glu Thr Leu Thr Phe Leu Leu Thr Gln Lys
65 70 75 80
Lys Thr Cys Val Lys Asn Tyr Val Gln Lys Glu
85 90
40

*Sub
a1
cont*